PERSON DETECTION BASED ON FUSION HISTOGRAM OF GRADIENTS WITH TEXTURE (FHGT) LOCAL FEATURES

Htet Htet Lin[1]

[1](Assistant Lecturer, Software Department, Computer University (BanMaw), Myanmar, htethtet.linnnnn@gmail.com)

Abstract— *Proficient real time objects detection is complicated and still active areas in computer vision due to many challenges: object appearance variations, intra-class and inter-class differences, difference articulation, illumination, static/ dynamic occlusions, and aspect variations. Extract discriminative and accurate features is also challenging in order to precise statistical data on monitoring people that allows users to make strategic decisions. In the previous works, they have achieved remarkable development for some people body parts, but less performances for all the full body. To tackle this appearance issue, this paper proposes to merge kth gradient differential with tamura texture features together to get discriminative and robust features. Specifically, the k order features are used to extract by differentiating and then combined these feature easily due to their same cell based feature to form the G features. Moreover statistical tamura texture features are extracted by using Gray Level Difference Matrix detector. Then, the system introduces a new insight powerful local image features, FHGT is proposed to capture the stronger object edges and shading variations, and local coherence of object appearance. Then, features combine by Joint Histograms. The experimental result is tested on the public Pascal VOC 2007 Dataset and results are outperformed.*

Keywords—People Detection; Kth Differential Gradients Features; Tamura Texture Features; FHGT

1. INTRODUCTION

People detection is always a hot research topic in computer vision (such as intelligent vehicles, robots, visual surveillance systems, operational, traffic and safety monitoring processes), which enters a fast-developing period especially after 2005. Generally, extracting feature from image regions have two distinct views on how to compute feature vectors: sparse or dense representation. The former approach is based on feature extraction from a salient image regions set. All these regions are not contain useful information: most are identical, texture less, or too cluttered to use. The later one densely focus on feature vectors of image regions. The new insight is that at the former' visual scene analysis, all areas might be identically essentials and it is not fail small details. For the later stages, they decide which regions are the most suitable to compute as region of interest. There have not any strong class specific assumptions but are less invariant to appearance changes and high miss detection rate [1].

In computer vision applications, texture feature extraction and analysis plays an important part. Specially, they are the four main texture types: (1) Statistical texture: They can be achieved the pixel of image' gray value by using the statistical approach. (2) Structural texture: They can be attained the gray pixel values by constructing some structural primitives. (3) Model based texture: They can be calculated to build model that based on the correlation of pixel gray value and its neighboring pixel gray value. (4) Transform based texture: They can be transformed with some techniques and using with statistical measurements [2].

HOG feature is widely used among the various types of feature due to its discriminative feature properties. Dalal and Triggs [3] studied the feature set of object detection, then they discovered the performance advantage of HOG for pedestrian detection via a series of contrast experiments between different feature descriptors about edge and gradient. They could be enhanced by combining with other types of features due to self-bias. Our finding is that a uniform model is not perfected for human detection due to fact of various appearances.

The detection rate is improved 7% by joining HOG with local binary pattern (LBP) [4]. However, the more the accuracy increased, the higher the computational costs. Inspired by some simple inherent attributes of people (i.e. appearance and shape), the proposed system aims to establish an efficient novel features for people detection system which can invariant to small variations in pose or appearance with accurate performance. So, this can be used in real-time application of surveillance, robotics and vehicle system.

According to the literature, many systems various issues still remain in numerous style of clothing in appearance, size, various variations and occluding. So, the proposed system will be developed as a novel framework for handling the appearance problems which can support better performance and more accurate detection rate. The entire operation is computed on cells-base and focused on overlapping blocks to avoid redundancy. The main contribution is highlighted a new insight, FHGT feature that based on the dense grids of locally normalized and overlapping encoding image regions to get the most significant and robustness features.

In this paper, the effective terminology of FHGT feature has been addressed. This paper also shows the distinct progress in detection process that has been achieved by the proposed methods. This paper organizes as follows. Section 2 presents the earlier methods and technologies and their applicable work of the detection system. Section 3 demonstrates the proposed approach to contribute in feature extraction field. Section 4 investigates and

discusses the experimental performance results compared with state-of-art results. The rest section, Section 5 has been attempted the conclusion and concerns.

2. RELATED WORK

In the last few decades, some technologies for detecting people was developed. Human detection based on feature is one of the most popular types of surveillance monitoring system. According to the literature, the previous works of human detection are surveyed in Fig. 1. This figure show the topology of human detection techniques. There are various methods of detecting people in accordance with the different between the feature selection and the classifier design. Either the system can single descriptor of feature extraction or can combine these functions for superior performance. There have been several studies related to descriptors of single attributes and combined feature descriptors.

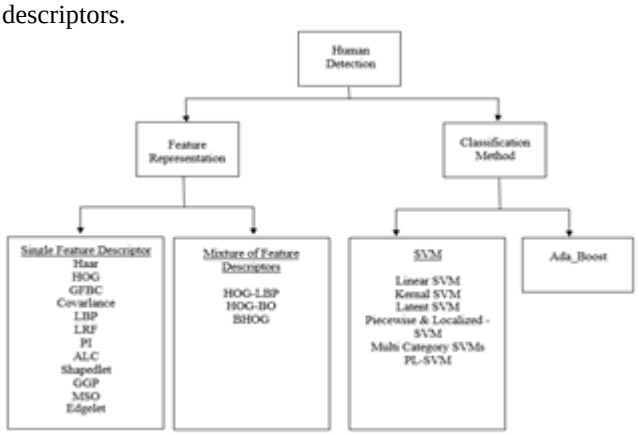

Fig. 1. Person Detection Methods

A. *Single Feature Descriptors for Person Detection Methods*

Descriptor of the HOG feature with linear support vector machine are proposed for people detection [5]. They can grasp the edges or a gradient structure. But, they can't handle the noisy marginal areas with exceptionally high computational cost. They only handle the image with a fixed scale ratio. Q. Zhu et.al [6] developed a fast HOG detection system using variable-size blocks with AdaBoost. Key of this method is 70 times faster than HOG [5]. It can handle multi-scale images. The weakness is their high computation cost. [7] introduced the recognition system of the objects with a cortex mechanism with SVM. They can identify objects on the background of debris and can handle multi scale images. But, their processing speed is very high, tens of seconds. And the computation cost is also high.

Yu. et al. [8] developed the semantic LBP and Fourier LBP system with AdaBoost. LBP is pure and can avoid the outliers. It has a low complexity and multi-scale extensions. But it has lower speed than HO and high computational cost. P. Sabzmeydani et.al [9] presented the Shapelet feature using AdaBoost classifier. This system can capture more data than fixed sets of features by spreading the classification outcomes. In [10], the structure of human detection focused on the integrated motion with the information about the intensity of the image. They used a threshold filter and this strategy was launched about 4 frames per second. This false positive level is very low due to low resolution of images under unpleasant situations. But, the main drawback is the high level of training time.

Q. Ye et.al [11] presented the framework of two-step MSO feature classifier for fast people detection. They focused on the coarse functions by Adaboost and focused on fine features by SVM. The main interest point of this system is that the processing speed is 12.4 times faster than [5] but they can't handle any occlusion treatment. P. F. Felzenszwalb et.al [12] developed an object detection using discriminant learning partial model. This method is effective, accurate and appropriate for representation class of variable objects. But, they wrongly recognized due to the confusion in the classroom.

B. *Combined Feature Descriptors for Person Detection Methods*

X. Wang et al. [13] developed a combination of LBP and HOG with SVM. Main advantages of this this system is effective for treating partial occlusion and is very effective for general detection of objects. These system are well outperformed on the INRIA data set compare other. But, the weakness of these system' detector can't handle with the deformation portion of the object. In [14], the PL-SVM people detection approach with different opinions and poses by joining BO and HOG as two cascaded. The detection of people with multiple views and multiple positions can be possible in these approach. This made the benefits of these system. PL-SVM has higher accuracy than linear, core and [5]. It operated on a cluttered background and can suppress noise. But, the time complexity is higher than the linear svm and difficult to cope with deformation of the object.

3. PROPOSED APPROACH

This paper describes the proposed framework of the feature detection with the intra-class and inter-class problem, where input is an image and then the output is the detection result. The quality measure of the output is based on the proposed feature. In the experiments, the proposed system improves the average precision rate than state-of-art approaches. This makes to capture our concerns.

The simplest way to detect edges in an image consists of identifying those points with a great change in the intensity gradient of the image. As an extension of the HOG features [3, 6], the proposed system augmented with kth order and introduced tamura features respectively. These resultant features are denoted as FHGT features. The system overview of the proposed approach is depicted in Fig. 2. In this figure, the original image is set as the input image of the proposed system. Then, the block including cells are taken to extract G feature and T feature. These two are combined by joint histogram approach. The proposed FHGT feature operating process consists of three main portions: G feature extracting, T feature extracting and combining these two features as shown in the algorithm, Fig. 3. And, then output result is out to detect the image.

A. *G Feature*

Preprocessing stage is also required to acquire the frame images and to convert HSI color space. Then, Hue and Saturation are combined to get the zero order feature. Kth order gradients features is differentiated from the (k – 1) th

order gradients value. Firstly, first order gradients is calculated, like HOG [3]. The second order gradients is very valuable for human detection. Second order is got according to the K order differential equations.

Both these features are easily combined together according to their cell based structure. Unlike [6], our system focused most cell in the overlapping blocks in the detecting window need to remove the redundant operations. It expressively increases the whole operation time of people detection.

B. T Feature

Based on the psychology study of human visual perception about texture, a new theory to express texture feature, Tamura. It focus on the probability estimation on the input' image gray level. The six components of texture correspond to the texture's six properties from psychological perspective, namely coarseness, contrast, directionality, lifelikeness, regularity, and roughness by using Grey Level Difference Matrix. This feature can show the image' size and the texture prominent elements level. Moreover, they extract the texture features directionality.

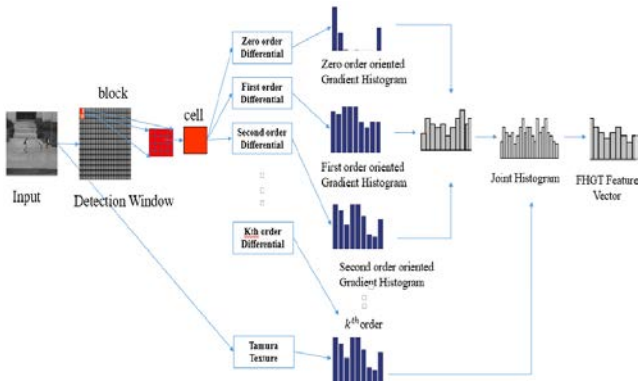

Fig. 2. System Overview of the Proposed System

C. Combining G and T Feature

The T and G features are combined by Joint Histogram to form FHGT features. The whole operations process are also show in Fig. 1. A joint histogram is especially critical in large images or combining many features of their histograms, because many images have the same histogram nature as shown in Fig. 4. It uses to merge the k^{th} gradients features (g feature) with cell based structures and t features. It decreases the time complexity. It accomplishes through careful selection of a set of local features for efficiency and robustness by reducing the space. It incorporates additional information from the image without sacrificing the robustness of histograms. Each entry in a joint histogram contains the number of pixels in the image that are described by a particular combination of feature values.

D. Classification process of human detection

FHGT features for the whole body of the human are classified by using Support Vector Machine (SVM) model that isolate regions of interest (features of the whole body) and judge to achieve real-time people detection. Using the created descriptor with a trained SVM allows determining the presence of a specific object, such as a human being. This classifier is binary and looks for the optimal hyper plane that maximizes the margin (gap) between the object class and non-object class in either the input feature space.

Test and train set of images are taken from Pascal dataset. The different styles, scale and position of the human can found in this images. So, the proposed system used this images to produce the classification model. In this paper, three testing process included with SVM. First part is the detection part of G feature extraction is tested with SVM. Second part is the detection part of T feature with SVM. And third part is used two SVMs are processed together like as [14], one for G feature and other one for T feature. If the input image is classified as human in former step G, T and FHGT feature will be extracted and tested with SVM in next stage.

A pair of observations are made from the testing operation over various images. To reduce the miss detection rate, the proposed system can enhance the initial model by running it over the false training set exhaustively.

```
Algorithm: Overall process of the proposed work for FGHT feature detection
Input: Images from the dataset
Output: Images with detection rate
If input = 0 then
        Error
Else
        A = preprocessing (input);
        B = gFeatureExtract (A);
        C = tFeatureExtract (A);
        D = dimensionMatch (B, C);
        E = jointHistogram (D);
        Output = Classify (E);
End
```

Fig. 3. Pseudo Code Representation of the Proposed System

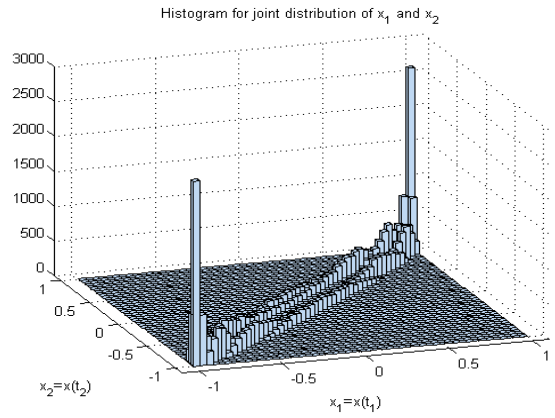

Fig. 4. Joint Histogram

4. EVALUATION PERFORMANCE

The evaluation are tested on Pascal Voc 2007 dataset. In extracting G feature, train-val set are used to train features. Note that the number of k^{th}gradients order is used two. But we will contribute k number is three as a future work. The overall performance accuracy and detection score densely depends on both how finely the test image is scanned and how the detections are merged. Table 1 show the full result of average precision tested for only person class. k^{th} Order features have not the same dimension. When their

dimensions are adjusted to get the same dimensions, 1.2% is improved.

The average precisions of T feature also show in Table 1. The whole result after combining G and T features show the improved performance than the state of art [6]. These experimental result is only for features extraction and feature combining. We will evaluate these features on different models as a future work.

TABLE I. COMPARISON OF DIFFERENT APPROACHES ON PASCAL VOC 2007 DATASET (AP IS AVERAGE PRECISION)

Feature	AP %
HSC [18]	44.14
HOG [3]	45.8
HOG III [17]	51.3
G without same dimensions	50.1
G	51.3
T	38.0
FHGT	52.1

5. CONCLUSION

A robust system for person detection by using combination of propose features (FHGT) is intended to solve the appearances problem. Firstly, extracting the edge contour and gradients features from these salient object by calculating the k^{th}order differential. Then, to improve the detection accuracy. These algorithm not only fuses multiple detections at nearby locations and scales and incorporates detection confidences, but also allows the detection of overlapping object instances occurring at very different scales. Experimental result are outperformed than the start of art methods.

REFERENCES

[1] N. Dalal. Finding people in images and videos. Institute National Poly technique de Grenoble-INPG, 2006.

[2] R. Filho, et al. Analysis of human tissue densities: A new approach to extract features from medical images. Pattern Recognition Letters, 2017.

[3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, in Proc. CVPR, 2005.

[4] M. Farhadi, S. A. Motamedi, and S. Sharifian, “Efficient human detection based on parallel implementation of gradient and texture feature extraction methods”, in Proc. IMVIP, 2011.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Jun. 2005, pp. 886–893.

[6] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, “Fast human detection using a cascade of histograms of oriented gradients,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Jul. 2006, pp. 1491–1498.

[7] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Object recognition with cortex-like mechanisms,” in Proc. IEEE Transaction on Pattern Analyzing and Machine Intelligence, vol. 29, no. 3, pp. 411–426, Mar. 2007.

[8] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, “Discriminative local binary patterns for human detection in personal album,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Jun. 2008, pp. 1–8.

[9] P. Sabzmeydani and G. Mori, “Detecting pedestrians by learning shapelet features,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Jun. 2007, pp. 1–8.

[10] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” International Journal of Computer Vision, vol. 63, no. 2, pp. 153–161, 2005.

[11] Q. Ye, J. Jiao, and B. Zhang, “Fast pedestrian detection with multi-scale orientation features and two-stage classifiers,” in Proc. IEEE 17th International Conference on Image Processing, Sep. 2010, pp. 881–884.

[12] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part based models,” IEEE Transaction on Pattern Analyzing and Machine Intelligent, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[13] X. Wang, T. X. Han, and S. Yan, “An HOG-LBP human detector with partial occlusion handling,” in Proc. IEEE International Conference Computer Vision, Oct. 2009, pp. 32–39.

[14] Q. Ye, Z. Han, J. Jiao, and J. Liu, “Human Detection in Images via Piecewise Linear Support Vector Machines,” in Proc. IEEE transactions on image processing, vol. 22, no. 2, February 2013.

[15] J. Stoble and S. Me, “Multi-posture human detection based on hybrid HOG-BO feature”, International Conference on Advances in Computing and Communications, 2015.

[16] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[17] Y. Jiang and J. Ma, “Combination Features and Models for Human Detection”, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[18] X. Ren and D. Ramanan, “Histograms of sparse codes for object detection”, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.